

Continuous information flow fluctuations

M.L. Rosinberg¹ and Jordan M. Horowitz²

¹*Laboratoire de Physique Théorique de la Matière Condensée,
Université Pierre et Marie Curie, CNRS UMR 7600,
4 place Jussieu, 75252 Paris Cedex 05, France**

²*Physics of Living Systems Group, Department of Physics,
Massachusetts Institute of Technology - 400 Technology Square, Cambridge, MA 02139*

Information plays a pivotal role in the thermodynamics of nonequilibrium processes with feedback. However, much remains to be learned about the nature of information fluctuations in small scale devices and their relation with fluctuations in other thermodynamics quantities, like heat and work. Here we derive a series of fluctuation theorems for information flow and partial entropy production in a Brownian particle model of feedback cooling and extend them to arbitrary driven diffusion processes. We then analyze the long-time behavior of the feedback-cooling model in detail. Our results provide insights into the structure and origin of large deviations of information and thermodynamic quantities in autonomous Maxwell's demons.

PACS numbers: 05.70.Ln, 05.40.-a, 89.70.-a

I. INTRODUCTION

Accounting for information is necessary to rationalize thermodynamics in the presence of feedback [1]. In small systems where noise is unavoidable, information not only bounds the average extracted work, but also fluctuates along individual stochastic trajectories, alongside heat and work. Although this feature has been incorporated in generalized (detailed and integral) fluctuation relations [2–9], little is known about the properties of information fluctuations and their correlations with other thermodynamic quantities. Of particular interest is the behavior at long times, an issue recently addressed for a discrete two-state information engine [10].

In this Letter, we analyze the large deviation statistics of information fluctuations for a Brownian particle model, which may be viewed as a dynamic version of a Maxwell's demon [1]. This model describes a feedback cooling (or cold damping) experiment [11] and has proven to be a rich playground for theoretical exploration [12–16]. We begin by proving a series of transient integral fluctuation theorems (IFTs). One of them, applicable to any coupled Langevin processes experiencing independent noises, is the analog of the IFT for bipartite Markov jump processes derived in [8]. We thereby confirm that we have identified the correct fluctuating analog of information flow in diffusion processes. With this groundwork, we calculate analytically and numerically the information flow fluctuations in the steady-state regime and their correlations with heat. This analysis reveals strong correlations that extend beyond the average behavior into the large, rare fluctuations regime. Further insight is then gained by unraveling the atypical trajectories that lead to such rare information or heat fluctuations.

II. SETUP

Consider a one-dimensional underdamped Brownian particle of mass m immersed in a thermal environment with viscous damping γ and temperature T . Feedback cooling is implemented by measuring the particle's velocity v_t using a low-pass filter with cut-off frequency $1/\tau$, and feeding back the measurement outcome y_t with gain κ as an additional friction force $f_t = -\kappa y_t$. The resulting dynamical evolution, including measurement and control, is captured by the coupled Langevin equations [16]

$$\begin{aligned} m\dot{v}_t &= -\gamma v_t - \kappa y_t + \xi_t \\ \tau \dot{y}_t &= -(y_t - v_t - \eta_t), \end{aligned} \quad (1)$$

where ξ_t is Gaussian thermal noise with zero mean and variance $\langle \xi_t \xi_{t'} \rangle = 2\gamma T \delta(t - t')$, and η_t is Gaussian measurement noise with zero mean and variance $\langle \eta_t \eta_{t'} \rangle = \Delta \delta(t - t')$. Here and throughout Boltzmann's constant is set to unity. An equivalent implementation of (1) in an electric circuit is discussed in [17] (see also [15] for a more general version of the model with a harmonic potential trapping the Brownian particle).

The feedback's purpose is to maintain the system in a nonequilibrium steady state (NESS), where the average kinetic temperature $T_{\text{kin}} \equiv m \langle v^2 \rangle_{\text{st}}$ is smaller than T (the subscript “st” indicates that the average is taken in the stationary state of the joint process). Consequently, the feedback controller must be extracting energy from a single heat reservoir and converting it into work, in apparent violation of the second law. However, the controller acts as a Maxwell's demon that autonomously gathers information in order to implement the feedback. This information saves the second law by providing a rigorous bound on the extracted work through any of several second-law-like inequalities, each utilizing a different notion of information [16]. We here focus on the “information flow” [18, 19] (or learning rate [20, 21]) and its fluctuations.

*Electronic address: mlr@lptmc.jussieu.fr

III. INFORMATION FLOW ON THE TRAJECTORY LEVEL

Stochastic information flow is a trajectory-level quantity that captures the dynamic variation of the correlations between y_t and v_t . To define this information flow, we require the time-dependent probability density $p_t(v, y)$, which evolves according to the Fokker-Planck equation [16]

$$d_t p_t(v, y) = -\partial_v J_t^v(v, y) - \partial_y J_t^y(v, y), \quad (2)$$

with probability currents

$$J_t^v(v, y) = -\frac{1}{m}(\gamma v + \kappa y)p_t(v, y) - \frac{\gamma T}{m^2}\partial_v p_t(v, y) \quad (3)$$

$$J_t^y(v, y) = -\frac{1}{\tau}(y - v)p_t(v, y) - \frac{\Delta}{2\tau^2}\partial_y p_t(v, y). \quad (4)$$

Consequently, the time derivative of the marginals is $d_t p_t(v) = -\partial_v J_t^v(v)$ with $J_t^v(v) = \int dy J_t^v(v, y)$, and similarly for $p_t(y)$.

We are here interested in the information flow due to the v -fluctuations, whose average enters the generalized second-law inequality [16]. This quantity is defined as the partial rate of change of the stochastic mutual information

$$I_t(v_t : y_t) = \ln \frac{p_t(v_t, y_t)}{p_t(v_t)p_t(y_t)}. \quad (5)$$

Namely, we split the total time variation of I_t as

$$\frac{dI_t}{dt} = -(\dot{v}_t^v + \dot{v}_t^y), \quad (6)$$

into a piece arising due to v -fluctuations

$$\begin{aligned} \dot{v}_t^v(v_t, y_t) &\equiv \frac{1}{p_t(v_t, y_t)} \partial_v J_t^v(v, y)|_{v_t, y_t} - \dot{v}_t \partial_v \ln p_t(v, y)|_{v_t, y_t} \\ &\quad - \frac{1}{p_t(v_t)} \partial_v J_t^v(v)|_{v_t} + \dot{v}_t \partial_v \ln p_t(v)|_{v_t} \end{aligned} \quad (7)$$

and a similar piece due to y -fluctuations, $\dot{v}_t^y(v_t, y_t)$. In these expressions, the probabilities and currents obtained by solving (2), are evaluated along the stochastic trajectories $\mathbf{v}_0^t \equiv \{v_{t'}\}_{0 \leq t' \leq t}$ and $\mathbf{y}_0^t \equiv \{y_{t'}\}_{0 \leq t' \leq t}$ generated by (1).

Like the stochastic Shannon entropy [22], \dot{v}_t^v has a mixed character as it depends on both the micro-state (v_t, y_t) and the whole ensemble of trajectories with initial density $p_0(v, y)$. The ensemble average correctly yields the information flow [16, 18, 19]: $\langle \dot{v}_t^v \rangle = -\int dv dy J_t^v(v, y) \partial_v \ln p_t(y|v)$. (Note that we choose a minus sign in (6) so that $\langle \dot{v}_t^v \rangle_{st} > 0$, like in [16]; the average information flow has an opposite sign in [18, 19].) By integrating $\dot{v}_t^v(v_t, y_t)$ over the time interval $[0, t]$, we introduce the trajectory observable, the integrated information current

$$I^v \equiv \int_0^t dt' \dot{v}_{t'}^v(v_{t'}, y_{t'}) = \int_0^t dt' \dot{s}_{t'}^v(v_{t'}, y_{t'}) + \ln \frac{p_t(v_t)}{p_0(v_0)}, \quad (8)$$

where $\dot{s}_t^v(v_t, y_t) = \partial_v J_t^v(v_t, y_t)/p_t(v_t, y_t) - \dot{v}_t \partial_v \ln p_t(v_t, y_t)$ is the time-variation of the stochastic joint entropy $s_t(v_t, y_t) = -\ln p_t(v_t, y_t)$ due to v 's dynamics.

IV. FLUCTUATION THEOREMS

The introduction of the time-integrated information current I^v allows us to generalize the integral fluctuation theorem (IFT) for the stochastic entropy production in the presence of continuous feedback. For an observer unaware of the existence of the controller, the apparent “total” entropy production in the time interval $[0, t]$ would be

$$\Sigma^v \equiv \Delta s - \frac{\mathcal{Q}}{T}, \quad (9)$$

where $\Delta s = -\ln p_t(v_t) + \ln p_0(v_0)$ is the entropy change in the system [22], and $\mathcal{Q} = \int_0^t dt' (-\gamma v_{t'} + \xi_{t'}) \circ v_{t'}$ is the heat received by the particle from the bath [23], corresponding to an entropy change in the medium $\Sigma^m = -\mathcal{Q}/T$ (the symbol \circ denotes a Stratonovich integral). However, as a result of the feedback, Σ^v is negative on average in the stationary cooling regime, and more generally does not verify a fluctuation theorem, $\langle e^{-\Sigma^v} \rangle \neq 1$. Missing in the exponential is the entropy change provided by the feedback mechanism. Guided by the case of Markov jump processes, we introduce a new observable, dubbed a “partial” entropy production [8],

$$\Sigma \equiv \Sigma^v + I^v = \Sigma^m + \int_0^t dt' \dot{s}_{t'}^v(v_{t'}, y_{t'}), \quad (10)$$

and demonstrate that it obeys the generalized IFT

$$\langle e^{-\Sigma} \rangle = 1, \quad (11)$$

where $\langle \dots \rangle$ denotes an average over all possible paths of duration t with initial state drawn from a distribution $p_0(v_0, y_0)$. This is a central result of this Letter. By Jensen's inequality, one recovers the second-law-like inequality $\langle \Sigma^v + I^v \rangle \geq 0$. In particular, the work extracted in the steady state obeys $\langle W_{\text{ext}} \rangle_{st} = -T \langle \Sigma^m \rangle_{st} \leq T \langle I^v \rangle_{st}$ (as $\langle \Delta s \rangle_{st} = 0$ and thus $\langle \Sigma^v \rangle_{st} = \langle \Sigma^m \rangle_{st}$).

Although Σ is not a coarse-grained observable since it is a functional of both trajectories \mathbf{v}_0^t and \mathbf{y}_0^t , it is non-trivial that it satisfies an IFT. To prove this result, we will show that Σ can be cast as the log-ratio of the probability $\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]$ of observing the trajectory $(\mathbf{v}_0^t, \mathbf{y}_0^t)$ to the probability $\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t]$ of observing the time-reversed trajectory $(\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t) = (-\mathbf{v}_t^0, \mathbf{y}_t^0)$ in a suitable defined *modified* dynamics:

$$\Sigma = \ln \frac{\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]}{\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t]}. \quad (12)$$

The modified dynamics only alters the measurement process and is generated by

$$\tau \dot{y}_t = y_t + v_t + \frac{\Delta}{\tau} \partial_y \ln p_t(-v_t, y_t) + \eta_t. \quad (13)$$

This dynamics is intimately related to the *auxiliary* or *driven* process [24–26] that generates the constrained path ensemble $\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t | \Sigma = \sigma t]$ at long times (see (S53) below).

The proof of (12) proceeds by first writing out (12) as

$$\Sigma = \ln \frac{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0] \mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, v_0] p_0(v_0, y_0)}{\mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0] \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)}, \quad (14)$$

where the path probabilities $\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0]$, $\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0]$, etc. are expressed in terms of Onsager-Machlup action functionals [27]. For instance,

$$\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0] \propto e^{-\frac{t}{2\tau}} e^{-\frac{1}{2\Delta} \int_0^t dt' [\tau \dot{y}_{t'} + y_{t'} - v_{t'}]^2}, \quad (15)$$

using the Stratonovich discretization. (As stressed in [16], the path functionals in (14) are not true conditional probabilities because v_t and y_t influence each other.) The conclusion follows by using the local detailed balance relation $\ln(\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0] / \mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]) = \Sigma^m$ and noting that the dynamics generated by (13) is absolutely continuous with respect to (1)

$$\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] = \mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, v_0] e^{\int_0^t dt' [\frac{d}{dt} s_{t'}(v_{t'}, y_{t'}) - \dot{s}_{t'}(v_{t'}, y_{t'})]}. \quad (16)$$

Details can be found in the Supplemental Material (SM). As Σ is a log-ratio of path probabilities, one readily obtains the IFT (11).

Note that the dynamics of v_t does not play any role in this calculation and that the force acting on y_t could be arbitrary. In other words, the IFT holds for any coupled Langevin processes involving independent noises [28] (as it holds for any bipartite Markov jump processes [8]). Consider for instance the two coupled overdamped Langevin equations

$$\begin{aligned} \dot{x}_t &= \mu_x F_1(x_t, y_t, t) + \xi_t^x \\ \dot{y}_t &= \mu_y F_2(x_t, y_t, t) + \xi_t^y, \end{aligned} \quad (17)$$

where $\langle \xi_t^i \xi_{t'}^j \rangle = 2D_i \delta_{ij} \delta(t - t')$ and $D_i = T_i \mu_i$. Then, replacing $F_2(x_t, y_t, t)$ by

$$F_2^*(x, y, t) = -F_2(x_t, y_t, t) + 2T_y \partial_y \ln p_t(x_t, y_t), \quad (18)$$

one can derive an IFT like (11) by a similar argument (see SM). Of course, a similar IFT holds for the partial entropy production of the y_t degree of freedom.

Two other fluctuation theorems are obtained in a similar manner by comparing the original feedback process to other modified dynamics. First, by flipping the sign of the viscous damping $\gamma \rightarrow -\gamma$ in the first Langevin equation,

$$m\dot{v}_t = \gamma v_t - \kappa y_t + \xi_t, \quad (19)$$

we obtain an IFT for the dissipated heat (or medium entropy production Σ^m)

$$\left\langle e^{-\Sigma^m} \right\rangle = e^{\frac{\gamma}{m} t}. \quad (20)$$

This relation, originally derived in [29], holds for any underdamped Langevin dynamics, provided the damping is linear. Finally, by combining the two dynamics (13) and (19), we find an IFT that includes information (see SM)

$$\left\langle \frac{p_t(v_t)}{p_0(v_0)} e^{-I^v} \right\rangle = e^{\frac{\gamma}{m} t}. \quad (21)$$

V. FLUCTUATIONS IN THE NESS

We now turn to calculating the stationary-state fluctuations of the integrated information current I^v and the medium entropy flow (or heat) Σ^m in our linear feedback cooling model (1). To simplify the notation, we drop the subscript “st”.

Thanks to the linearity of (1), the stationary distribution of the joint system is Gaussian

$$p(v, y) = \frac{1}{\sqrt{(2\pi^2)|\mathbf{C}|}} e^{-\frac{1}{2}(v, y) \cdot \mathbf{C}^{-1} \cdot (v, y)^T}, \quad (22)$$

where the entries of the covariance matrix \mathbf{C} ($c_{11} = \langle v^2 \rangle$, $c_{12} = \langle vy \rangle$, $c_{22} = \langle y^2 \rangle$) are given in Appendix D of [16]. From (7) and (8), we then obtain the explicit expression for I^v :

$$\begin{aligned} I^v &= -A_1 t + \frac{1}{2}(\alpha_{11} - \frac{1}{\sigma_{11}})(v_t^2 - v_0^2) \\ &+ \int_0^t dt' (\alpha_{11} A_1 v_{t'}^2 + A_2 y_{t'}^2 + A_3 v_{t'} y_{t'} + \alpha_{12} \dot{v}_{t'} y_{t'}), \end{aligned} \quad (23)$$

where $\alpha_{11} = c_{22}/|\mathbf{C}|$, $\alpha_{12} = -c_{12}/|\mathbf{C}|$, and

$$\begin{aligned} A_1 &= \frac{\gamma}{m} (1 - T \frac{\alpha_{11}}{m}), \quad A_2 = \frac{\alpha_{12}}{m} (\kappa - \frac{\gamma T}{m} \alpha_{12}) \\ A_3 &= \frac{a}{m} \alpha_{11} + \frac{\gamma}{m} \alpha_{12} (1 - 2T \frac{\alpha_{11}}{m}). \end{aligned} \quad (24)$$

A. Large deviation analysis

As $t \rightarrow \infty$, the stationary probability distribution of a time-integrated observable \mathcal{A} – such as I^v or Σ^m – is said to satisfy a large deviation principle if it takes the scaling form $P(\mathcal{A} = at) \sim e^{-tE(a)}$, where $E(a)$ is the large deviation rate function (LDF) [30]. As usual, it is convenient to introduce the associated moment generating function $Z_a(\lambda, t) = \langle e^{-\lambda \mathcal{A}} \rangle$, which behaves asymptotically as

$$Z_a(\lambda, t) \sim g_a(\lambda) e^{t\mu_a(\lambda)}, \quad (25)$$

where $\mu_a(\lambda) \equiv \lim_{t \rightarrow \infty} (1/t) \ln Z_a(\lambda, t)$ is the scaled cumulant generating function (SCGF) and $g_a(\lambda)$ is a sub-leading factor. The LDF is normally obtained via the Legendre transform $E(a) = -[\mu_a(\lambda^*(a)) + \lambda^*(a)a]$, with

the saddle point λ^* determined by $\mu'_a(\lambda^*(a)) = -a$ [30]. This relation, however, breaks down if $g_a(\lambda)$ has a singularity in the region of the saddle-point integration, due to rare but large fluctuations of a boundary temporal term (*e.g.* the second term in the first line of (23)). The leading contribution to the LDF then comes from the singularity, which induces an exponential tail in the pdf. As stressed in [29], this may even make the SCGF discontinuous at $\lambda = 1$ when the modified process, such as the one governed by (13), has no stationary density.

We calculate the SCGFs for the medium entropy production rate $\sigma^m = \Sigma^m/t$ and the information flow $i^v = I^v/t$ by direct integration of the path probability. The calculation is made tractable by imposing periodic boundary conditions on the trajectories, which allows us to expand v_t and y_t in a discrete Fourier series (see *e.g.* [31, 32]). The results are conveniently expressed in terms of the response function in the frequency domain

$$\chi(\omega) = \frac{1 - i\omega\tau}{\kappa + \gamma - i(m + \gamma\tau)\omega - m\omega^2\tau}, \quad (26)$$

and two auxiliary functions

$$F_{\sigma^m, \lambda}(\omega) = \frac{2\gamma\Delta\kappa^2}{T} \frac{|\chi(\omega)|^2}{1 + \omega^2\tau^2} \left(1 - \frac{2T}{\Delta\kappa} - \lambda\right) \lambda \quad (27a)$$

$$F_{i^v, \lambda}(\omega) = 2\gamma T \Delta \alpha_{12}^2 \frac{|\chi(\omega)|^2}{1 + \omega^2\tau^2} \left[\left(\frac{\kappa}{m} \frac{\alpha_{11}}{\alpha_{12}} - \frac{\gamma}{m}\right)^2 + \omega^2 \right] \lambda^2, \quad (27b)$$

as (see SM)

$$\mu_{\sigma^m}(\lambda) = - \int_0^\infty \frac{d\omega}{2\pi} \ln[1 - F_{\sigma^m, \lambda}(\omega)] \quad (28a)$$

$$\mu_{i^v}(\lambda) = A_1 \lambda - \int_0^\infty \frac{d\omega}{2\pi} \ln[1 - F_{i^v, \lambda}(\omega)]. \quad (28b)$$

We highlight two important features of (27)-(28). First, $\mu_{\sigma^m}(\lambda)$ has the symmetry $\mu_{\sigma^m}(\lambda) = \mu_{\sigma^m}(1 - \frac{2T}{\Delta\kappa} - \lambda)$, which implies that the pdf $P(\Sigma^m)$ satisfies the steady-state fluctuation theorem

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{P(\Sigma^m = \sigma^m t)}{P(\Sigma^m = -\sigma^m t)} = \left(1 - \frac{2T}{\Delta\kappa}\right) \sigma^m, \quad (29)$$

at least in a limited range of σ^m around 0 (see Fig. 2 below). Second, $F_{i^v, \lambda}(\omega)$ is an even function of λ . Therefore, the LDF $E(i^v)$ is symmetric around the expectation value $\langle i^v \rangle = -\mu'_{i^v}(0) = -A_1$. We will elaborate on this point below. (On the other hand, the SCGF of Σ , given by a similar but more complicated expression - see SM -, does not display any symmetry.)

To study the correlations between Σ^m and I^v in the long-time limit, we also compute by the same method the SCGF $\mu_{\sigma^m, i^v}(\lambda_1, \lambda_2) = \lim_{t \rightarrow \infty} (1/t) \ln \langle e^{-\lambda_1 \Sigma^m - \lambda_2 I^v} \rangle$ and the corresponding joint LDF $E(\sigma^m, i^v)$ (see SM). As an added benefit, this allows us to investigate the fluctuations of the ratio $\epsilon = -\Sigma^m/I^v$ that characterizes the efficiency of the information-to-work conversion along the trajectories. The most probable value of ϵ is the “macroscopic” efficiency $\bar{\epsilon} = -\langle \sigma^m \rangle / \langle i^v \rangle$ [5, 33, 34].

B. Numerical study

To further explore the fluctuations of trajectory observables, we now present some numerical results. Hereafter the model is described by three dimensionless parameters: the feedback gain $g = \kappa/\gamma$, the signal-to-noise ratio $\text{SNR} = 2T/(\gamma\Delta)$, and the ratio τ/τ^v where $\tau^v = m/\gamma$ is the velocity relaxation time. Specifically, we set $\tau/\tau^v = 0.01$ and $\text{SNR} = 40$ and vary the feedback gain g , as in experiments [11]. By choosing a moderate noise level, we make the control more sensitive to information-flow fluctuations [35].

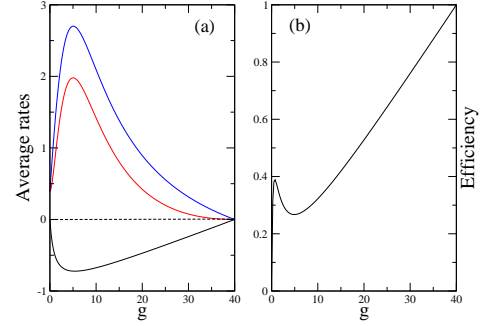


FIG. 1: (Color on line) (a) The average rates $\langle \sigma^m \rangle$ (black line), $\langle \sigma \rangle$ (red line), and $\langle i^v \rangle$ (blue line) as a function of the feedback gain g . One has $\langle \sigma \rangle = \langle \sigma^m \rangle + \langle i^v \rangle \geq 0$. (b) The most probable efficiency $\bar{\epsilon} = -\langle \sigma^m \rangle / \langle i^v \rangle$ as a function of g .

Let us first recall that cooling [$T_{\text{kin}} < T$ and thus $\langle \sigma^m \rangle = (1/\tau^v)(T_{\text{kin}} - T) < 0$] requires $g/\text{SNR} < 1$, independent of the value of τ (cf. eq. (20) in [16] with the noise variance σ^2 replaced by Δ). This is illustrated in Fig. 1(a) where we plot the average rates as a function of g . The most salient features are the extrema in $\langle \sigma^m \rangle$ and $\langle i^v \rangle$. The minimum in $\langle \sigma^m \rangle$ occurs at $g = g_{\text{opt}} = \sqrt{1 + \text{SNR}} - 1$ [16]. Above g_{opt} , too much measurement noise is fed back to the system, causing T_{kin} to increase with g , a well-known experimental fact [11]. Eventually, for $g/\text{SNR} > 1$, the system is heated instead of cooled. On the other hand, for $\tau \neq 0$, the maximum in the information flow occurs at $g_{\text{KB}} = g_{\text{opt}}[1 - (\tau/\tau^v)\sqrt{1 + \text{SNR}}] < g_{\text{opt}}$. The demon then realizes a *Kalman-Bucy* filter [36] (hence the notation g_{KB}). In this limit, $\hat{v}_t \equiv (\tau/\tau^v)g_{\text{opt}}y_t$ represents the best estimate of v_t in terms of the mean-squared error $\mathcal{E}_t = \langle (v_t - \hat{v}_t)^2 \rangle$, given all past measurements [16, 17] (recall that (1) describes a non-Markovian control protocol [4]). Interestingly, as shown in [16], $\langle i^v \rangle$ is then equal to the transfer entropy rate $g_{\text{opt}}/2$. We see in Fig. 1(b) that these extrema in $\langle \sigma^m \rangle$ and $\langle i^v \rangle$ induce a local minimum in the information efficiency $\bar{\epsilon}$. This minimum would exactly occur at g_{opt} if τ were zero (note that $\bar{\epsilon} = 1$ for $g/\text{SNR} = 1$, when the demon does not extract any work.).

We now fix the gain at its optimal value g_{opt} and inves-

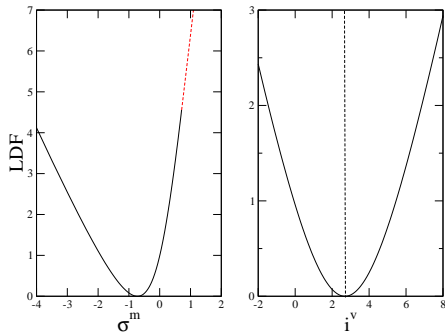


FIG. 2: (Color on line) Large deviation functions $E(\sigma^m)$ (left) and $E(i^v)$ (right) at maximum power ($g = g_{\text{opt}}$). The FT symmetry $E(\sigma^m) - E(-\sigma^m) = (\frac{2T}{\Delta\kappa} - 1)\sigma^m$ is obeyed for $|\sigma^m| < 0.712$ whereas $E(\sigma^m)$ is linear for $\sigma^m > 0.712$ (dashed red line). $E(i^v)$ is symmetric around its minimum.

tigate the fluctuations. We stress that g_{opt} is close to g_{KB} with our choice of the parameters, so that $\langle i^v \rangle$ is almost maximal (and the so-called sensory capacity [21] is close to 1). The large deviation functions $E(\sigma^m)$ and $E(i^v)$ are plotted in Fig. 2. Notice that $E(\sigma^m)$ has a linear branch for $\sigma^m > 0.712$ due to the presence of a pole in the pre-exponential factor. As a result, the steady-state FT (29) does not hold for large values of σ^m (this, however, corresponds to extremely rare events and depends on the choice of the model parameters). More intriguing is the symmetry exhibited by $E(i^v)$ around its minimum, which we already pointed out. This implies that positive and negative fluctuations of I^v around the expectation value are equiprobable. While we have no complete analytic proof nor heuristic argument, we believe this symmetry is a general property of our model even for finite-time fluctuations. This is indeed suggested by numerical simulations as well as by a small- t expansion of the modified generating function $Z_{i^v}(\lambda, t)e^{\lambda\langle i^v \rangle t} = \langle e^{-\lambda[I^v - \langle I^v \rangle]} \rangle$ displaying only even powers of λ (see SM). It should be added that perturbative calculations show that the symmetry is lost when there are nonlinearities in the feedback control [42].

The LDF for the efficiency ϵ is obtained from the joint SCGF $\mu_{\sigma^m, i^v}(\lambda_1, \lambda_2)$ as $E(\epsilon) = -\inf_{\lambda_1} \mu_{\sigma^m, i^v}(\lambda_1, \epsilon\lambda_1)$ and is plotted in Fig. 3. Like in the case of stochastic heat engines [37–41], $E(\epsilon)$ is a non-monotonic function with a minimum at the most probable value $\bar{\epsilon}$ and equal asymptotes for $\epsilon \rightarrow \pm\infty$ (the convergence to the asymptotic limit is slow, likely following a power law [39, 40]). We observe that the least probable value of ϵ , corresponding to the maximum of $E(\epsilon)$, is *negative*. This feature distinguishes the present “information engine” from the stochastic heat engines studied in [37–41].

Having determined the LDFs for information and heat, we now turn to the structure and origin of these fluctuations. We begin by addressing the typical information required to produce a rare fluctuation of heat, or the other

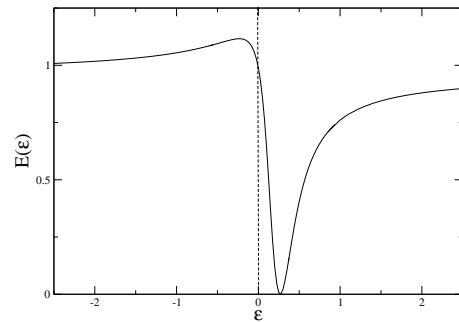


FIG. 3: Large deviation function of the efficiency $E(\epsilon)$ for $g = g_{\text{opt}}$. The most and least probable values are $\epsilon_{\text{most}} = \bar{\epsilon} \approx 0.268$ and $\epsilon_{\text{least}} \approx -0.225$.

way around. In Fig. 4, we plot the most probable value of i^v (resp. σ^m) for a given rare fluctuation of σ^m (resp. i^v). These quantities are computed from the derivatives of the joint SCGF $\mu_{\sigma^m, i^v}(\lambda_1, \lambda_2)$ at $\lambda_2 = 0$ (resp. $\lambda_1 = 0$) (see SM) (note that we restrict our study to the range $\sigma^m < 0.712$ where fluctuations of the boundary term are irrelevant).

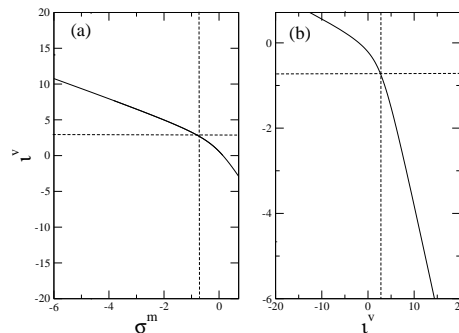


FIG. 4: (a) Most probable value of i^v for a given value of σ^m . (b) Most probable value of σ^m for a given value of i^v . The dashed lines indicate the typical values $\langle \sigma^m \rangle \approx -0.72$ and $\langle i^v \rangle \approx 2.70$.

As could be expected intuitively, the fluctuations of σ^m and i^v are strongly correlated. But a less predictable and remarkable feature is the asymmetry between positive and negative fluctuations: observing a negative fluctuation $\sigma^m < \langle \sigma^m \rangle$ (resp. a positive fluctuation $i^v > \langle i^v \rangle$) for a long time requires a smaller (resp. larger) variation of i^v (resp. σ^m) than observing $\sigma^m > \langle \sigma^m \rangle$ (resp. $i^v < \langle i^v \rangle$).

Deeper insight into the origin of these fluctuations is offered by studying the two auxiliary (or driven) processes [24–26] that generate the constrained ensembles $\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t | \Sigma^m = \sigma^m t]$ or $\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t | I^v = i^v t]$ asymptotically. Rare fluctuations then become typical. These auxiliary dynamics are again linear (see SM for details), with mod-

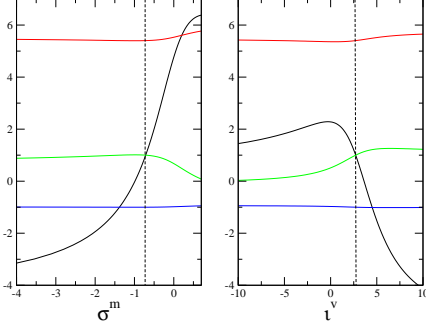


FIG. 5: (Color on line) Effective interactions of the auxiliary processes that generate atypical values of σ^m (left panel) or v^v (right panel): γ_{eff} (black line), κ_{eff} (red line), k_1 (blue line), k_2 (green line). Vertical dashed lines indicate the typical values of σ^m and v^v .

ified effective interactions

$$\begin{aligned} m\dot{v}_t &= -\gamma_{\text{eff}}v_t - \kappa_{\text{eff}}y_t + \xi_t \\ \tau\dot{y}_t &= k_1y_t + k_2v_t + \eta_t. \end{aligned} \quad (30)$$

The variations of the coefficients with σ^m or v^v are shown in Fig. 5. Again, we observe different behavior for negative and positive fluctuations. The atypical events $\sigma^m < \langle \sigma^m \rangle$ or $v^v > \langle v^v \rangle$ are created essentially by a decrease of the friction coefficient, which even becomes negative (κ_{eff} and k_2 also vary, but slightly, and k_1 does not change). As a result, the fluctuations of v_t are enhanced, as confirmed by Fig. 6, and the effective kinetic temperature increases. This additional uncertainty about v_t allows for more information to be gathered, leading to the corresponding increase in v^v observed in Fig. 3a. In other words, acquiring more information does not necessarily mean that it is effectively used to cool the system.

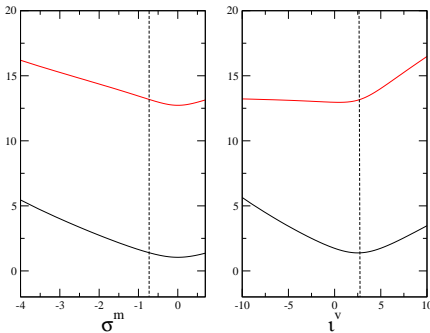


FIG. 6: (Color on line) $\langle v_t^2 \rangle$ (black line) and $\langle y_t^2 \rangle$ (red line) conditioned on a given value of σ^m (left) or v^v (right).

The case of atypical events $\sigma^m > \langle \sigma^m \rangle$ or $v^v < \langle v^v \rangle$ is not so simple, as both γ_{eff} and k_2 vary significantly.

Moreover, γ_{eff} is not a monotonic function of v^v . Remarkably, k_2 becomes very small as v^v becomes very negative. The demon then does not perform any measurement, and y_t just plays the role of additional noise.

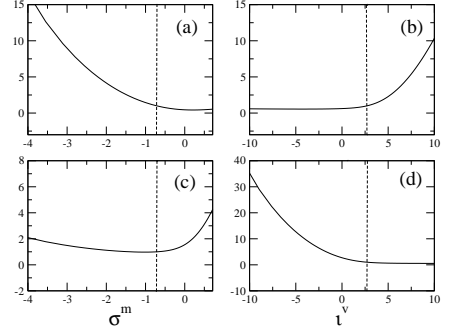


FIG. 7: Intensity of the atypical noises as a function of σ^m or v^v . (a) and (b): $S_{\xi_{\text{atyp}}}(\omega=0)/(2\gamma T)$; (c) and (d): $S_{\eta_{\text{atyp}}}(\omega=0)/\Delta$. Vertical dashed lines indicate the typical values of σ^m and v^v .

An alternative approach to understanding the origin of large deviations is to consider the atypical noise realizations that create rare fluctuations and determine whether fluctuations in ξ_t or η_t play the dominant role. To this end, we select an atypical trajectory $(\mathbf{v}_0^t, \mathbf{y}_0^t)_{\text{atyp}}$ produced by the auxiliary processes (S53) for a given value of σ^m or v^v and insert it into the original equations of motion (1). This yields two biased noises ξ_{atyp} and η_{atyp} . Thanks to the linearity of the equations, this calculation can be performed by working in the frequency domain, which yields

$$\begin{aligned} \xi_{\text{atyp}}(\omega) &= (\gamma - im\omega)v_{\text{atyp}}(\omega) + \kappa y_{\text{atyp}}(\omega) \\ &= \frac{\chi_{\text{eff}}(\omega)}{k_1 + i\omega\tau} \left[[(\gamma - im\omega)(k_1 + i\omega\tau) - \kappa k_2]\xi(\omega) \right. \\ &\quad \left. + [(\gamma - im\omega)\kappa_{\text{eff}} - (\gamma_{\text{eff}} - im\omega)\kappa]\eta(\omega) \right] \\ \eta_{\text{atyp}}(\omega) &= (1 - i\omega\tau)y_{\text{atyp}}(\omega) - v_{\text{atyp}}(\omega) \\ &= -\frac{\chi_{\text{eff}}(\omega)}{k_1 + i\omega\tau} \left[[(k_1 + i\omega\tau) + k_2(1 - i\omega\tau)]\xi(\omega) \right. \\ &\quad \left. + [(\gamma_{\text{eff}} - im\omega)(1 - i\omega\tau) + \kappa_{\text{eff}}]\eta(\omega) \right], \end{aligned} \quad (31)$$

with

$$\chi_{\text{eff}}(\omega) = \frac{k_1 + i\omega\tau}{(\gamma_{\text{eff}} - im\omega)(k_1 + i\omega\tau) - \kappa_{\text{eff}}k_2}. \quad (32)$$

Therefore, the two noises are correlated and colored, with power spectral densities $S_{\xi_{\text{atyp}}}(\omega) = \langle \xi_{\text{atyp}}(\omega)\xi_{\text{atyp}}(-\omega) \rangle$ and $S_{\eta_{\text{atyp}}}(\omega) = \langle \eta_{\text{atyp}}(\omega)\eta_{\text{atyp}}(-\omega) \rangle$. For simplicity, we here only characterize ξ_{atyp} and η_{atyp} by their intensity, that is the zero-frequency part of their power spectrum. The variations of $S_{\xi_{\text{atyp}}}(\omega=0)$ and $S_{\eta_{\text{atyp}}}(\omega=0)$ (normalized by the intensities of the original white noises) as a function of σ^m or v^v are shown in Fig. 7. The

overall picture is again very instructive: atypical events $\sigma^m < \langle \sigma^m \rangle$ or $v^v > \langle v^v \rangle$ are mainly due to an atypical history of the thermal noise ξ_t whereas atypical events $\sigma^m > \langle \sigma^m \rangle$ or $v^v < \langle v^v \rangle$ are mainly due to an atypical history of the measurement noise η_t .

VI. CONCLUSION

In this Letter, we have studied a Brownian particle model of feedback cooling, as a realization of an autonomous Maxwell's demon. We have first derived a series of fluctuation theorems for the information flow and the partial entropy production in coupled diffusion processes. We then investigated the fluctuations of information flow and entropy flow, and their correlations, focusing on the long-time limit. By analyzing in detail

the effective dynamics and atypical noise realizations that instigate rare fluctuations, we have unraveled the subtle trade-off between noise in the system, noise in the measurement device, and control efficiency. We believe that using the same approach with other models of Maxwell's demon could significantly improve our understanding of the thermodynamics of information.

Acknowledgments

We wish to thank the organizers of the workshop *New Frontiers in Non-equilibrium Physics* held at the Yukawa Institute of Theoretical Physics (Kyoto, 2015) where this work was initiated. MLR also thanks T. Munakata for helpful discussions. JMH is supported by the Gordon and Betty Moore Foundation through Grant GBMF4343.

-
- [1] PARRONDO J. M. R., HOROWITZ J. M. and SAGAWA T., *Nature Physics*, **11** (2015) 131.
 - [2] SAGAWA T. and UEDA M., *Phys. Rev. Lett.*, **104** (2010) 090602.
 - [3] HOROWITZ J. M. and VAIKUNTANATHAN S., *Phys. Rev. E*, **82** (2010) 061120.
 - [4] SAGAWA T. and UEDA M., *Phys. Rev. E*, **85** (2012) 021104.
 - [5] ESPOSITO M. and SCHALLER G., *Eur. Phys. Lett.*, **99** (2012) 30003.
 - [6] ITO S. and SAGAWA T., *Phys. Rev. Lett.*, **111** (2013) 180603.
 - [7] HARTICH D., BARATO A.C and SEIFERT U., *J. Stat. Mech.: Theory Exp.* (2014) P02016.
 - [8] SHIRAISHI N. and SAGAWA T., *Phys. Rev. E*, **91** (2015) 012130.
 - [9] KOSKI J. V., MAISI V. F., SAGAWA T. and PEKOLA J. P., *Phys. Rev. Lett.*, **113** (2014) 030601.
 - [10] MAITLAND M., GROSSKINSKY S. and HARRIS R. J., *Phys. Rev. E*, **92** (2015) 052136.
 - [11] For a review and reference therein, see e.g. POOT M. and VAN DER ZANT S. J., *Physics Reports*, **511** (2012) 273.
 - [12] KIM K. H. and QIAN H., *Phys. Rev. E*, **75** (2007) 022102.
 - [13] ITO S. and SANO S., *Phys. Rev.*, **84** (2011) 021123.
 - [14] MUNAKATA T. and ROSINBERG M. L., *J. Stat. Mech.: Theory Exp.* (2012) P05010.
 - [15] MUNAKATA T. and ROSINBERG M. L., *J. Stat. Mech.: Theory Exp.* (2013) P06014.
 - [16] HOROWITZ J. M. and SANDBERG H., *New Journal of Physics*, **16** (2014) 125007.
 - [17] SANDBERG H., DELVENNE J.-C., NEWTON N. J. and MITTER S. K., *Phys. Rev. E*, **90** (2014) 042119.
 - [18] ALLAHVERDYAN A. E., JANZING D. and MAHLER G., *J. Stat. Mech.: Theory Exp.* (2009) P09011.
 - [19] HOROWITZ J. M. and ESPOSITO M., *Phys. Rev. X*, **4** (2014) 031015.
 - [20] BARATO A. C., HARTICH D. and SEIFERT U., *New J. Phys.*, **16** (2014) 103024.
 - [21] HARTICH D., BARATO A. C. and SEIFERT U., *Phys. Rev. E*, **93** (2016) 022116.
 - [22] SEIFERT U., *Phys. Rev. Lett.*, **95** (2005) 040602.
 - [23] SEKIMOTO K., *Prog. Theor. Phys. Suppl.*, **130** (1998) 17.
 - [24] GARRAHAN J. P. and LESANOVSKY I., *Phys. Rev. Lett.*, **104** (2010) 160601.
 - [25] JACK R. L. and SOLLICH P., *Prog. Theor. Phys. Supp.*, **184** (2010) 304.
 - [26] CHETRIT R. and TOUCHETTE H., *Phys. Rev. Lett.*, **111** (2013) 120601; *Ann. Inst. Poincaré A*, **16** (2015) 2005.
 - [27] Note that the initial state distribution of the backward path is $p_t(v_t, y_t)$, the final state distribution of the forward path, and not $p_t(-v_t, y_t)$.
 - [28] HOROWITZ J. M., *J. Stat. Mech.: Theory Exp.* (2015) P03006.
 - [29] ROSINBERG M. L., TARJUS G. and MUNAKATA T., *Eur. Phys. Lett.*, **113** (2016) 10007.
 - [30] TOUCHETTE T., *Phys. Rep.*, **1** (2009) 478.
 - [31] ZAMPONI F., BONETTO F., CUGLIANDOLO L. F. and KURCHAN J., *J. Stat. Mech.: Theor. Exp.* (2005) P09013.
 - [32] KUNDU A., SABHAPANDIT S. and DHAR A., *J. Stat. Mech.: Theory Exp.* (2011) P03007.
 - [33] CAO F. J. and FEITO M., *Phys. Rev. E*, **79** (2009) 041118.
 - [34] BAUER M., ABREU D. and SEIFERT U., *J. Phys. A Math. Theor.*, **45** (2012) 162001.
 - [35] BECHHOEFER J., *New J. Phys.*, **17** (2015) 075003.
 - [36] ASTROM K. J. and MURRAY R. M. *Feedback systems: An introduction for scientists and engineers* Princeton University Press, New Jersey, 2008.
 - [37] VERLEY G., WILLAERT T., VAN DEN BROECK C. and ESPOSITO M., *Nat. Commun.* **5**, 4721 (2014); *Phys. Rev. E* **90**, 052145 (2014).
 - [38] GINGRICH T. R., ROTSKOFF G. M., VAIKUNTANATHAN S. and GEISSLER P. L., *New J. Phys.* **16**, 102003 (2014).
 - [39] POLETTINI M., VERLEY G., and ESPOSITO M., *Phys. Rev. Lett.* **114**, 050601 (2015).
 - [40] PROESMANS K., CLEUREN B. and VAN DEN

BROECK C., *Eur. Phys. Lett.*, **109** (2015) 20004.
 [41] MARTINEZ I. A., ROLDAN E., DINIS L., PETROV D.,
 PARRONDO J. M. R., and RICA R., *Nat. Phys.* **12**, 67
 (2016).

[42] On the other hand, the symmetry is preserved if one adds
 a force depending linearly on the position x_t of the Brownian
 particle in the first Langevin equation.

Supplemental Material

I. FLUCTUATION RELATIONS

A. IFT for Σ

We first derive the IFT for the (partial) entropy production $\Sigma = \Sigma^v + I^v$. As stated in the main text, this boils down to showing that

$$\Sigma = \ln \frac{\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]}{\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t]}, \quad (\text{S1})$$

where $\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t] \equiv \mathcal{P}^*[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0] \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)$ is the joint probability density of the time-reversed path $(\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t) = (-\mathbf{v}_t^0, \mathbf{y}_t^0)$ generated by a modified process, hereafter denoted by the star symbol. The probabilities $\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0]$, $\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0]$, etc. are expressed in terms of Onsager-Machlup (OM) action functionals.

Following [8], we anticipate that the star dynamics only modifies the equation of motion for y_t , so that $\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t] = \mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0] \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)$. Therefore,

$$\ln \frac{\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]}{\mathcal{P}^*[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t]} = \Sigma^m + \ln \frac{\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0] p_0(v_0, y_0)}{\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)}, \quad (\text{S2})$$

where we have used the local detailed balance relation $\ln(\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0] / \mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]) = \Sigma^m$. Introducing the quantity $\dot{s}_t^v(v_t, y_t)$, we see that eq. (S1) is satisfied if

$$\begin{aligned} \ln \frac{\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0]}{\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0]} &= \int_0^t dt' \dot{s}_{t'}^v(v_{t'}, y_{t'}) + \ln \frac{p_t(v_t, y_t)}{p_0(v_0, y_0)} \\ &\equiv - \int_0^t dt' \dot{s}_{t'}^y(v_{t'}, y_{t'}) \end{aligned} \quad (\text{S3})$$

where $\dot{s}_t^y(v_t, y_t) \equiv \frac{d}{dt} s_t(v_t, y_t) - \dot{s}_t^v(v_t, y_t)$ is the time-variation of the stochastic joint entropy $s_t(v_t, y_t)$ due to y 's dynamics. Hence

$$\dot{s}_t^y(v_t, y_t) = \frac{1}{p_t(v_t, y_t)} \partial_y J_t^y(v_t, y_t) - \dot{y}_t \partial_y \ln p_t(v_t, y_t). \quad (\text{S4})$$

Thus, the derivation of the IFT rests on the construction of a Langevin process that generates trajectories \mathbf{y}_0^t with a conditional weight $\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0]$ obeying eq. (S3). As we now show, this process is governed by the equation of motion

$$\tau \dot{y}_t = v_t + y_t + \frac{\Delta}{\tau} \partial_y \ln p_t(-v_t, y_t) + \eta_t, \quad (\text{S5})$$

where η_t is the same Gaussian white noise as in the original process. Equation (S5) may be viewed as an overdamped Langevin equation with a time-dependent force $F^*(v_t, y_t, t) = v_t + y_t + (\Delta/\tau) \partial_y \ln p_t(-v_t, y_t)$ and τ^{-1} playing the role of a mobility μ_y . Hence

$$\begin{aligned} \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] &\propto e^{-\frac{1}{2\Delta} \int_0^t dt' [\tau \dot{y}_{t'} - F^*(v_{t'}, y_{t'}, t')]^2 - \frac{1}{2\tau} \int_0^t dt' \partial_y F(v_{t'}, y_{t'}, t')} \\ &\propto e^{-\frac{t}{2\tau}} e^{-\frac{1}{2\Delta} \int_0^t dt' \left\{ [\tau \dot{y}_{t'} - v_{t'} + y_{t'} + \frac{\Delta}{\tau} \partial_y \ln p_{t'}(v_{t'}, y_{t'})]^2 + \left(\frac{\Delta}{\tau}\right)^2 \partial_y^2 \ln p_{t'}(v_{t'}, y_{t'}) \right\}}, \end{aligned} \quad (\text{S6})$$

where we have changed t' into $-t'$ in the second line of the equation (we recall that the velocity v_t is odd under time reversal but that y_t must be treated as an even variable [16]). Comparing with the conditional density associated with the original Langevin equation,

$$\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0] \propto e^{\frac{t}{2\tau}} e^{-\frac{1}{2\Delta} \int_0^t dt' [\tau \dot{y}_{t'} + y_{t'} - v_{t'}]^2}, \quad (\text{S7})$$

we obtain

$$\ln \frac{\mathcal{P}[\mathbf{y}_0^t | \mathbf{v}_0^t, y_0]}{\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0]} = \frac{t}{\tau} + \frac{1}{\tau} \int_0^t dt' \left\{ (\tau \dot{y}_{t'} + y_{t'} - v_{t'}) \partial_y \ln p_{t'}(v_{t'}, y_{t'}) + \frac{\Delta}{2\tau} \frac{\partial_y^2 p_{t'}(v_{t'}, y_{t'})}{p_{t'}(v_{t'}, y_{t'})} \right\}, \quad (\text{S8})$$

where we have used the identity $(\partial_y \ln f)^2 + \partial_y^2 \ln f = f^{-1} \partial_y^2 f$. On the other hand, by inserting the expression of the probability current $J_t^y(v_t, y_t) = -\tau^{-1}(y_t - v_t)p(v_t, y_t) - \Delta/(2\tau^2)\partial_y p_t(v_t, y_t)$ into eq. (S4), we find

$$\dot{s}_t^y(v_t, y_t) = -\frac{1}{\tau} - \frac{1}{\tau}(y_t - v_t)\partial_y \ln p_t(v_t, y_t) - \frac{\Delta}{2\tau^2} \frac{\partial_y^2 p_t(v_t, y_t)}{p_t(v_t, y_t)} - \dot{y}_t \partial_y \ln p_t(v_t, y_t). \quad (\text{S9})$$

Therefore eq. (S3) is satisfied, as announced.

As stated in the main text, the IFT holds for any coupled Langevin processes involving independent noises, for instance the two overdamped Langevin equations (17). Eq. (S8) is then replaced by

$$\begin{aligned} \ln \frac{\mathcal{P}[\mathbf{y}_0^t | \mathbf{x}_0^t, y_0]}{\mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{x}}_0^t, \tilde{y}_0]} &= \frac{1}{4T_y} \int_0^t dt' \left[2\dot{y}_{t'} - \mu_y [F_2(x_{t'}, y_{t'}, t') - F_2^*(x_{t'}, y_{t'}, t')] \right] \left[F_2(x_{t'}, y_{t'}, t') + F_2^*(x_{t'}, y_{t'}, t') \right] \\ &\quad - \frac{\mu_y}{2} \int_0^t dt' \partial_y [F_2(x_{t'}, y_{t'}, t') - F_2^*(x_{t'}, y_{t'}, t')], \end{aligned} \quad (\text{S10})$$

where $F_2^*(x, y, t)$ is the force acting on y_t in the modified dynamics. Then, by choosing

$$F_2^*(x, y, t) = -F_2(x, y, t) + 2T_y \partial_y \ln p_t(x, y), \quad (\text{S11})$$

the r.h.s. of eq. (S10) identifies with $\int_0^t dt' \dot{s}_t^y(x_{t'}, y_{t'})$, with $\dot{s}_t^y(v_t, y_t)$ defined by eq. (S4). We leave the demonstration as an exercise for the reader.

B. IFT for I^v

The IFT for the integrated information flow [eq. (21) in the main text] is obtained by modifying also the dynamics of v_t and changing γ into $-\gamma$ while keeping the variance of ξ_t fixed (this modified process is denoted by the hat symbol hereafter). As shown in [28], this leads to

$$\frac{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t, v_0]}{\hat{\mathcal{P}}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]} = e^{\frac{\gamma}{m}t} e^{\Sigma^m}. \quad (\text{S12})$$

By replacing the trajectories by their time-reversed images, this relation can be rewritten as

$$\Sigma^m = \frac{\gamma}{m}t + \ln \frac{\hat{\mathcal{P}}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]}{\mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]}. \quad (\text{S13})$$

Therefore, using eq. (S1), we have

$$\begin{aligned} \Sigma - \Sigma^m &= -\frac{\gamma}{m}t + \ln \frac{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t]}{\mathcal{P}^*[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t]} - \ln \frac{\hat{\mathcal{P}}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]}{\mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]} \\ &= -\frac{\gamma}{m}t + \ln \frac{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t]}{\mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0] \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)} - \ln \frac{\hat{\mathcal{P}}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]}{\mathcal{P}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0]} \\ &= -\frac{\gamma}{m}t + \ln \frac{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t]}{\hat{\mathcal{P}}^*[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t]}, \end{aligned} \quad (\text{S14})$$

where $\hat{\mathcal{P}}^*[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t] \equiv \hat{\mathcal{P}}[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t, \tilde{v}_0] \mathcal{P}^*[\tilde{\mathbf{y}}_0^t | \tilde{\mathbf{v}}_0^t, \tilde{y}_0] p_t(v_t, y_t)$ is the joint probability density of the backward path generated by the modified “hat-star” processes (i.e., the “hat” dynamics for v_t and the “star” dynamics for y_t). Since $\Sigma - \Sigma^m = I^v - \ln p_t(v_t)/p_0(v_0)$ (cf. eqs. (8) and (9) in the main text), we finally obtain

$$\frac{p_t(v_t)}{p_0(v_0)} e^{-I^v} = e^{\frac{\gamma}{m}t} \frac{\hat{\mathcal{P}}^*[\tilde{\mathbf{v}}_0^t | \tilde{\mathbf{y}}_0^t]}{\mathcal{P}[\mathbf{v}_0^t | \mathbf{y}_0^t]}, \quad (\text{S15})$$

which leads to the IFT for the information flow by integration over the ensemble of paths generated by the original process. Note that this IFT is less general than the one for Σ since the force acting on the state variable of the first subsystem must be linear. This is true for the model under study since this force is just the viscous force $-\gamma v_t$. For the coupled overdamped Langevin processes described by eqs. (17), one must have $F_1(x, y, t) = -kx_t + F(y_t, t)$. Then $(\gamma/m)t$ is replaced by $(\mu_x k)t$ in eqs. (S12)-(S15) (see [28] for a more general discussion).

II. SCALED CUMULANT GENERATING FUNCTIONS

In this section we derive eqs. (27) in the main text, and we discuss the domain of validity of these expressions. We also give the expressions of $\mu_\sigma(\lambda)$ and of the joint SCGF $\mu_{\sigma^m, \iota^v}(\lambda_1, \lambda_2)$. These results are obtained by imposing periodic boundary conditions on the trajectories and expanding v_t and y_t in discrete Fourier series,

$$\begin{aligned} v(t) &= \sum_{n=-\infty}^{\infty} v_n e^{-i\omega_n t} \\ y(t) &= \sum_{n=-\infty}^{\infty} y_n e^{-i\omega_n t}, \end{aligned} \quad (\text{S16})$$

with inverse transforms

$$\begin{aligned} v_n &= \frac{1}{t} \int_0^t ds v(s) e^{i\omega_n s} \\ y_n &= \frac{1}{t} \int_0^t ds y(s) e^{i\omega_n s}, \end{aligned} \quad (\text{S17})$$

where $\omega_n = 2\pi n/t$ and $v_n \equiv v(\omega_n)$, $y_n \equiv y(\omega_n)$. The coupled Langevin equations in the frequency domain are then

$$\begin{aligned} (\gamma - im\omega_n)v_n &= -\kappa y_n + \xi_n \\ (1 - i\omega_n\tau)y_n &= v_n + \eta_n, \end{aligned} \quad (\text{S18})$$

with $\langle \xi_n \xi_{n'} \rangle = (2\gamma T/t)\delta_{n, -n'}$, $\langle \eta_n \eta_{n'} \rangle = (\Delta/t)\delta_{n, -n'}$, and $\langle \xi_n \eta_{n'} \rangle = 0$. This leads to

$$\begin{aligned} v_n &= \chi_n \left[\xi_n - \frac{\kappa}{1 - i\omega_n\tau} \eta_n \right] \\ y_n &= \frac{\chi_n}{1 - i\omega_n\tau} [\xi_n + (\gamma - im\omega_n)\eta_n] \end{aligned} \quad (\text{S19})$$

where

$$\chi_n \equiv \chi(\omega_n) = \frac{1 - i\omega_n\tau}{\kappa + \gamma - i(m + \gamma\tau)\omega_n - m\omega_n^2\tau}. \quad (\text{S20})$$

A. SCGFs for single observables

For brevity, we will only detail the calculation of $\mu_{\iota^v}(\lambda)$. In the stationary state, I^v is given by eq. (23) in the main text. We recall that α_{11}, α_{22} and $\alpha_{12} = \alpha_{21}$ are the entries of \mathbf{C}^{-1} , the inverse of the covariance matrix whose expression is given in Appendix D of [16]. In terms of the Fourier coefficients v_n and y_n , the rate $\iota^v \equiv I^v/t$ reads

$$\iota^v = -A_1 + \sum_{n=-\infty}^{\infty} [\alpha_{11}A_1v_nv_{-n} + A_2y_ny_{-n} + (A_3 - i\omega_n\alpha_{12})v_ny_{-n}] + b.t. \quad (\text{S21})$$

where $b.t.$ is a temporal boundary term that is neglected hereafter (see the discussion below). Inserting eqs. (S19) then yields

$$\begin{aligned} \iota^v &\sim -A_1 + \sum_{n=-\infty}^{\infty} \frac{|\chi_n|^2}{1 + \omega_n^2\tau^2} \left\{ [\alpha_{11}A_1(1 + \omega_n^2\tau^2) + A_2 + (A_3 - i\omega_n\alpha_{12})(1 - i\omega_n\tau)]|\xi_n|^2 \right. \\ &\quad + [\kappa^2\alpha_{11}A_1 + (\gamma^2 + m^2\omega_n^2)A_2 - \kappa(A_3 - i\omega_n\alpha_{12})(\gamma + im\omega_n)]|\eta_n|^2 \\ &\quad + [-\kappa\alpha_{11}A_1(1 - i\omega_n\tau) + (\gamma + im\omega_n)A_2 + (A_3 - i\omega_n\alpha_{12})(\gamma + im\omega_n)(1 - i\omega_n\tau)]\xi_n\eta_{-n} \\ &\quad \left. + [-\kappa\alpha_{11}A_1(1 + i\omega_n\tau) + (\gamma - im\omega_n)A_2 - \kappa(A_3 - i\omega_n\alpha_{12})]\xi_{-n}\eta_n \right\}, \end{aligned} \quad (\text{S22})$$

which is rewritten in a compact form as

$$\iota^v \sim -A_1 + \sum_{n=1}^{\infty} \zeta_n^T \mathbf{L}_{n, \iota^v} \zeta_n^*, \quad (\text{S23})$$

where $\zeta_n = (\xi_n, \eta_n)^T$, the symbol $*$ denotes the complex conjugate (with $\xi_n^* = \xi_{-n}$, $\eta_n^* = \eta_{-n}$), and $\mathbf{L}_{n,iv}$ is a 2×2 Hermitian matrix with entries

$$\begin{aligned} L_{11,iv}(\omega_n) &= \frac{2|\chi_n|^2}{1 + \omega_n^2 \tau^2} [\alpha_{11} A_1 (1 + \omega_n^2 \tau^2) + A_2 + A_3 - \omega_n^2 \tau \alpha_{12}] \\ L_{12,iv}(\omega_n) &= L_{21}^*(\omega_n) = \frac{|\chi_n|^2}{1 + \omega_n^2 \tau^2} \left[-2\kappa \alpha_{11} A_1 (1 - i\omega_n \tau) + 2(\gamma + im\omega_n) A_2 + A_3 [(\gamma + im\omega_n)(1 - i\omega_n \tau) - \kappa] \right. \\ &\quad \left. - i\omega_n \alpha_{12} [(\gamma + im\omega_n)(1 - i\omega_n \tau) + \kappa] \right] \\ L_{22,iv}(\omega_n) &= \frac{2|\chi_n|^2}{1 + \omega_n^2 \tau^2} [\kappa^2 \alpha_{11} A_1 + (\gamma^2 + m^2 \omega_n^2) A_2 - \kappa (A_3 \gamma + m\omega_n^2 \alpha_{12})] . \end{aligned} \quad (\text{S24})$$

This leads to

$$\langle e^{-\lambda I^v} \rangle \sim e^{\lambda A_1 t} \prod_{n=1}^{\infty} \int d\zeta_n P(\zeta_n) e^{-\lambda t \zeta_n^T \mathbf{L}_{n,iv} \zeta_n^*} , \quad (\text{S25})$$

where

$$P(\zeta_n) = \frac{1}{\pi^2 \det \mathbf{D}} e^{-\zeta_n \mathbf{D}^{-1} \zeta_n^*} \quad (\text{S26})$$

and

$$\mathbf{D} = \frac{1}{t} \begin{pmatrix} 2\gamma T & 0 \\ 0 & \Delta \end{pmatrix} .$$

The Gaussian integration over ζ_n then gives

$$\int d\zeta_n P(\zeta_n) e^{-\lambda t \zeta_n^T \mathbf{L}_{n,iv} \zeta_n^*} = \det[\mathbf{I} + \lambda t \mathbf{D} \mathbf{L}_{n,iv}]^{-1} . \quad (\text{S27})$$

In the long-time limit, the summation over n can be replaced by an integral over ω , and we finally obtain

$$\mu_{iv}(\lambda) = \frac{\gamma}{m} (1 - T \frac{\alpha_{11}}{m}) \lambda - \int_0^\infty \frac{d\omega}{2\pi} \ln[1 - F_{iv,\lambda}(\omega)] , \quad (\text{S28})$$

with

$$F_{iv,\lambda}(\omega) = -2\gamma T \Delta \lambda^2 [L_{11,iv}(\omega) L_{22,iv}(\omega) - L_{12,iv}(\omega) L_{21,iv}(\omega)] , \quad (\text{S29})$$

which leads to eq. (27b) in the main text.

The calculation of $\mu_{\sigma^m}(\lambda)$ is quite similar but somewhat simpler since $\Sigma^m = -(\kappa/T) \int_0^t dt' v_{t'} y_{t'} \sim -(\kappa/T) \sum_n v_n y_n$. The entries of the corresponding matrix \mathbf{L}_{n,σ^m} are then given by

$$\begin{aligned} L_{11,\sigma^m}(\omega_n) &= -\frac{2\kappa}{T} \frac{|\chi_n|^2}{1 + \omega_n^2 \tau^2} \\ L_{12,\sigma^m}(\omega_n) &= L_{21}^*(\omega_n) = -\frac{\kappa}{T} \frac{|\chi_n|^2}{1 + \omega_n^2 \tau^2} [(\gamma + im\omega_n)(1 - i\omega_n \tau) - \kappa] \\ L_{22,\sigma^m}(\omega_n) &= \frac{2\kappa^2 \gamma}{T} \frac{|\chi_n|^2}{1 + \omega_n^2 \tau^2} , \end{aligned} \quad (\text{S30})$$

so that

$$\mu_{\sigma^m}(\lambda) = - \int_0^\infty \frac{d\omega}{2\pi} \ln[1 - F_{\sigma^m,\lambda}(\omega)] , \quad (\text{S31})$$

with

$$F_{\sigma^m,\lambda}(\omega) = -\lambda [2\gamma T L_{11,\sigma^m}(\omega) + \Delta L_{22,\sigma^m}(\omega)] - 2\gamma T \Delta \lambda^2 [L_{11,\sigma^m}(\omega) L_{22,\sigma^m}(\omega) - L_{12,\sigma^m}(\omega) L_{21,\sigma^m}(\omega)] , \quad (\text{S32})$$

which leads to eq. (27a) in the main text. The main difference with $F_{i^v,\lambda}(\omega)$ is that the term linear in λ does not vanish so that $F_{\sigma^m,\lambda}(\omega)$ is not an even function of λ .

Finally, the expression of $\mu_\sigma(\lambda)$ is obtained by exploiting the fact that $\Sigma \sim \Sigma^m + I^v$ as $t \rightarrow \infty$. Hence

$$\mu_\sigma(\lambda) = \frac{\gamma}{m}(1 - T\frac{\alpha_{11}}{m})\lambda - \int_0^\infty \frac{d\omega}{2\pi} \ln[1 - F_{\sigma,\lambda}(\omega)] , \quad (\text{S33})$$

with

$$F_{\sigma,\lambda}(\omega) = -\lambda[2\gamma T(L_{11,\sigma}(\omega) + \Delta L_{22,\sigma}(\omega))] - 2\gamma T\Delta\lambda^2[L_{11,\sigma}(\omega)L_{22,\sigma}(\omega) - L_{12,\sigma}(\omega)L_{21,\sigma}(\omega)] \quad (\text{S34})$$

and $L_{ij,\sigma} = L_{ij,\sigma^m} + L_{ij,i^v}$.

Note that the SCGFs are real quantities in an open domain (λ_-, λ_+) (different for each function) in which the argument of the logarithm stays positive for all values of ω . For instance, with the choice $\tau/\tau^v = 0.01$, $\text{SNR} = 40$, $g = g_{\text{opt}} \approx 5.403$ for the model parameters, we find that $\mu_{i^v}(\lambda)$, $\mu_{\sigma^m}(\lambda)$, and $\mu_\sigma(\lambda)$ are defined in the intervals $-1.004 < \lambda < 1.004$, $-8.130 < \lambda < 1.727$, and $-1.931 < \lambda < 1.027$, respectively. The slopes of the SCGFs diverge at the boundaries, which implies that the corresponding Legendre transforms are asymptotically linear [29]. However, this is no longer true if the pre-exponential factors (see eq. (25) in the main text and eq. (S44) below) have pole singularities inside the domain of definition. These singularities result from rare but large fluctuations of the boundary terms neglected in the preceding calculation, and the leading contribution to the LDF then comes from the singularity (whose position fixes the slope of the LDF). For instance, $g_{\sigma^m}(\lambda)$ diverges for $\lambda = \lambda_0 \approx -6.39$, so that $E(\sigma^m) \approx 0.038 + 6.39\sigma^m$ for $\sigma^m > -\mu'_{\sigma^m}(\lambda_0) \approx 0.712$ (see Fig. 2 of the main text).

B. Joint SCGFs

The same method based on discrete Fourier transforms can be used to compute joint SCGFs such as $\mu_{\sigma^m,i^v}(\lambda_1, \lambda_2) \equiv \lim_{t \rightarrow \infty} (1/t) \ln \langle e^{-\lambda_1 \Sigma^m - \lambda_2 I^v} \rangle$. Moreover, since $\Sigma \sim \Sigma^m + I^v$ in the long-time limit, the three functions $\mu_{\sigma^m,i^v}(\lambda_1, \lambda_2)$, $\mu_{\sigma^m,\sigma}(\lambda_1, \lambda_2)$, $\mu_{\sigma,i^v}(\lambda_1, \lambda_2)$ are not independent. Specifically, $\mu_{\sigma,i^v}(\lambda_1, \lambda_2) = \mu_{\sigma^m,i^v}(\lambda_1, \lambda_1 + \lambda_2)$ and $\mu_{\sigma^m,\sigma}(\lambda_1, \lambda_2) = \mu_{\sigma^m,i^v}(\lambda_1 + \lambda_2, \lambda_2)$, with

$$\mu_{\sigma^m,i^v}(\lambda_1, \lambda_2) = \frac{\gamma}{m}(1 - T\frac{\alpha_{11}}{m})\lambda_2 - \frac{1}{2} \int_{-\infty}^\infty \frac{d\omega}{2\pi} \ln[1 - F_{\sigma^m,i^v,\lambda_1,\lambda_2}(\omega)] \quad (\text{S35})$$

and

$$\begin{aligned} F_{\sigma^m,i^v,\lambda_1,\lambda_2}(\omega) = & -2\gamma T[\lambda_1 L_{11,\sigma^m}(\omega) + \lambda_2 L_{11,i^v}(\omega)] - \Delta[\lambda_1 L_{22,\sigma^m}(\omega) + \lambda_2 L_{22,i^v}(\omega)] \\ & - 2\gamma T\Delta \left\{ [\lambda_1 L_{11,\sigma^m}(\omega) + \lambda_2 L_{11,i^v}(\omega)][\lambda_1 L_{22,\sigma^m}(\omega) + \lambda_2 L_{22,i^v}(\omega)] \right. \\ & \left. - [\lambda_1 L_{12,\sigma^m}(\omega) + \lambda_2 L_{12,i^v}(\omega)][\lambda_1 L_{12,\sigma^m}^*(\omega) + \lambda_2 L_{12,i^v}^*(\omega)] \right\} . \end{aligned} \quad (\text{S36})$$

One can readily check that $\mu_{\sigma^m}(\lambda) = \mu_{\sigma^m,i^v}(\lambda, 0)$, $\mu_\sigma(\lambda) = \mu_{\sigma^m,i^v}(0, \lambda)$, and $\mu_\sigma(\lambda) = \mu_{\sigma^m,i^v}(\lambda, \lambda)$.

III. TILTED AND AUXILIARY PROCESSES

In this section, we construct the so-called auxiliary (or driven) processes [24-26] that describe how large fluctuations of Σ^m , I^v , or Σ are created in the long-time limit. For each of these observables, we first determine the so-called tilted generator and compute the dominant eigenvalue and the associated left and right eigenfunctions (the dominant eigenvalue identifies with the SCGF already obtained in section II A, but the knowledge of the eigenfunctions allows us to compute the pre-exponential factors). We then determine the biased forces or biased noises that make a large deviation of the observable typical.

For brevity, we mostly focus on large deviations of I^v . We also closely follow the analysis and notations of [26].

A. Spectral elements

1. Tilted generators

From the definition of I^v [eqs. (7) and (8) in the main text], we note that this quantity belongs to the general class of trajectory observables of the form [26]

$$I^v = \int_0^t f_{i^v}(\mathbf{X}_t) dt' + \int_0^t \mathbf{g}_{i^v}(\mathbf{X}_{t'}) \circ d\mathbf{X}_{t'} , \quad (\text{S37})$$

where $\mathbf{X}_t = (v_t, y_t)$, $f_{i^v}(v, y) = \partial_v J^v(v, y)/p(v, y)$ (since $\partial_v J(v) = 0$ in the stationary state), and $\mathbf{g}_{i^v}(v, y)$ is a two-dimensional vector function with components $(-\partial_v \ln p(y|v), 0)$. We then introduce the non-conservative process associated with the exponentially tilted trajectory ensemble

$$\mathcal{P}_{i^v, \lambda}[\mathbf{v}_0^t, \mathbf{y}_0^t] \equiv \frac{e^{-\lambda I^v} \mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]}{\langle e^{-\lambda I^v} \rangle} \quad (\text{S38})$$

that becomes equivalent in the limit $t \rightarrow \infty$ to the ensemble of trajectories conditioned on a particular value of i^v (the equivalence holds because the LDF $E(i^v)$ is convex, and the value of λ achieving the equivalence is then given by $\lambda = -E'(i^v)$). The corresponding tilted generator is given by [26]

$$\mathcal{L}_{i^v, \lambda} = \mathbf{F}(\nabla - \lambda \mathbf{g}_{i^v}) + (\nabla - \lambda \mathbf{g}_{i^v}) \frac{\mathbf{D}'}{2} (\nabla - \lambda \mathbf{g}_{i^v}) - \lambda f_{i^v} , \quad (\text{S39})$$

where \mathbf{F} is the two-dimensional drift of the original process with components $(-(\gamma v + \kappa y)/m, (v - y)/\tau)$ and the noise covariance matrix \mathbf{D}' is here defined as

$$\mathbf{D}' = \begin{pmatrix} \frac{2\gamma T}{m^2} & 0 \\ 0 & \frac{\Delta}{\tau^2} \end{pmatrix} .$$

The generating function $Z_{i^v, \lambda}(v, y, t) = \langle e^{-\lambda I^v} \delta(v - v_t) \delta(y - y_t) \rangle$ (where the average is taken over all trajectories of duration t ending at (v, y) with initial state (v_0, y_0) drawn from the stationary pdf $p(v_0, y_0)$) evolves according to

$$\partial_t Z_{i^v, \lambda}(v, y, t) = \mathcal{L}_{i^v, \lambda}^\dagger Z_{i^v, \lambda}(v, y, t) , \quad (\text{S40})$$

where $\mathcal{L}_{i^v, \lambda}^\dagger$ is the dual of $\mathcal{L}_{i^v, \lambda}$. Namely,

$$\begin{aligned} \mathcal{L}_{i^v, \lambda}^\dagger = & \frac{\gamma T}{m^2} \frac{\partial^2}{\partial v^2} + \frac{\Delta}{2\tau^2} \frac{\partial^2}{\partial y^2} + \frac{1}{m} \left[\gamma v + \kappa y - 2 \frac{\lambda \gamma T}{m} \partial_v \ln p(y|v) \right] \partial_v + \frac{y - v}{\tau} \partial_y \\ & + \frac{\lambda}{m} \left[\gamma + (\gamma v + \kappa y) \partial_v \ln p(v) \right] + \frac{\lambda \gamma T}{m^2} \left[(\lambda + 1) (\partial_v \ln p(y|v))^2 + 2 \partial_v \ln p(y|v) \partial_v \ln p(v) \right. \\ & \left. + (\partial_v \ln p(v))^2 + \frac{\partial^2}{\partial v^2} \ln p(v) \right] + \frac{\gamma}{m} + \frac{1}{\tau} , \end{aligned} \quad (\text{S41})$$

with $\partial_v \ln p(y|v) = -(\alpha_{11} - c_{11}^{-1})v - \alpha_{12}y$ and $\partial_v \ln p(v) = -c_{11}^{-1}v$ (we recall that c_{ij} and α_{ij} are the entries of the covariance matrix and its inverse, respectively).

In order to determine the asymptotic behavior of $Z_{i^v, \lambda}(v, y, t)$, we do not need to solve the spectral problem for $\mathcal{L}_{i^v, \lambda}$ and $\mathcal{L}_{i^v, \lambda}^\dagger$ in full generality but only to compute the dominant eigenvalue $\mu_{i^v}(\lambda)$ and the associated right and left eigenfunctions, solutions of the equations

$$\begin{aligned} \mathcal{L}_{i^v, \lambda} r_{i^v, \lambda}(v, y) &= \mu_{i^v}(\lambda) r_{i^v, \lambda}(v, y) \\ \mathcal{L}_{i^v, \lambda}^\dagger l_{i^v, \lambda}(v, y) &= \mu_{i^v}(\lambda) l_{i^v, \lambda}(v, y) , \end{aligned} \quad (\text{S42})$$

with normalization conditions $\int dv dy l_{i^v, \lambda}(v, y) = 1$ and $\int dv dy r_{i^v, \lambda}(v, y) l_{i^v, \lambda}(v, y) = 1$. From eq. (S41) and the corresponding expression of $\mathcal{L}_{i^v, \lambda}$, we find that the bivariate Gaussian functions $r_{i^v, \lambda}(v, y) = e^{-(1/2)[A_{i^v}(\lambda)v^2 + B_{i^v}(\lambda)y^2 + 2C_{i^v}(\lambda)vy]}$ and $l_{i^v, \lambda}(v, y) = e^{-(1/2)[A_{i^v}^\dagger(\lambda)v^2 + B_{i^v}^\dagger(\lambda)y^2 + 2C_{i^v}^\dagger(\lambda)vy]}$ are solutions (not yet normalized) of these equations, with the coefficients $A_{i^v}(\lambda)$, $A_{i^v}^\dagger(\lambda)$, etc. obeying complicated algebraic equations. $\mu_{i^v}(\lambda)$ is eventually obtained as the solution of an algebraic equation of degree 6, and the proper root is selected by imposing

numerical agreement with eq. (S28). Then, assuming the existence of a gap between μ_{i^v} and the first sub-dominant eigenvalue, $Z_{i^v,\lambda}(v, y, t)$ takes the asymptotic form [26]

$$Z_{i^v,\lambda}(v, y, t) \sim e^{\mu_{i^v}(\lambda)t} \int dv_0 dy_0 p(v_0, y_0) r_{i^v,\lambda}(v_0, y_0) l_{i^v,\lambda}(v, y) , \quad (\text{S43})$$

so that $Z_{i^v}(\lambda, t) = \int dv dy Z_{i^v,\lambda}(v, y, t) \sim g_{i^v}(\lambda) e^{\mu_{i^v}(\lambda)t}$ with

$$g_{i^v}(\lambda) = \int dv_0 dy_0 p(v_0, y_0) r_{i^v,\lambda}(v_0, y_0) . \quad (\text{S44})$$

Depending on the model parameters (e.g. the feedback gain), the above integral may diverge for certain values of λ , signaling that the fluctuations of the temporal boundary terms must not be neglected, as discussed at the end of section B.1.

Similar calculations are performed for Σ^m and Σ . For completeness, we report the corresponding expressions of the tilted dual generators $\mathcal{L}_{\sigma^m,\lambda}^\dagger$ and $\mathcal{L}_{\sigma,\lambda}^\dagger$:

$$\begin{aligned} \mathcal{L}_{\sigma^m,\lambda}^\dagger &= \frac{\gamma T}{m^2} \frac{\partial^2}{\partial v^2} + \frac{\Delta}{2\tau^2} \frac{\partial^2}{\partial y^2} + \frac{1}{m} [(1-2\lambda)\gamma v + \kappa y] \partial_v + \frac{y-v}{\tau} \partial_y \\ &\quad + \lambda \left[(\lambda-1)\gamma \frac{v^2}{T} - \frac{\gamma}{m} \right] + \frac{\gamma}{m} + \frac{1}{\tau} , \end{aligned} \quad (\text{S45})$$

and

$$\begin{aligned} \mathcal{L}_{\sigma,\lambda}^\dagger &= \frac{\gamma T}{m^2} \frac{\partial^2}{\partial v^2} + \frac{\Delta}{2\tau^2} \frac{\partial^2}{\partial y^2} + \frac{1}{m} \left[(1-2\lambda)\gamma v + \kappa y - 2\frac{\lambda\gamma T}{m} \partial_v \ln p(v, y) \right] \partial_v + \frac{y-v}{\tau} \partial_y \\ &\quad + \lambda \left[\gamma v [(\lambda-1)\frac{v}{T} + \frac{2\lambda}{m} \partial_v \ln p(v, y)] + (\lambda+1) \frac{\gamma T}{m^2} (\partial_v \ln p(v, y))^2 \right] + \frac{\gamma}{m} + \frac{1}{\tau} . \end{aligned} \quad (\text{S46})$$

Note that $\mathcal{L}_{\sigma^m,\lambda=1}^\dagger - \mathcal{L}_{\sigma^m,\lambda=0}^\dagger = -2(\gamma/m)v\partial_v - (\gamma/m)$ so that $\partial_t Z_{\sigma^m}(1, t) = (\gamma/m)Z_{\sigma^m}(1, t)$ by integration over v and y , in agreement with the IFT $Z_{\sigma^m}(1, t) = \langle e^{-\Sigma^m} \rangle = e^{(\gamma/m)t}$. Likewise, $\mathcal{L}_{\sigma,\lambda=1}^\dagger - \mathcal{L}_{\sigma,\lambda=0}^\dagger = 2(\gamma/m)v + (T/m)\partial_v \ln p(v, y)[\partial_v \ln p(v, y) - \partial_v]$ so that $p(v, y)$ is an eigenfunction of $\mathcal{L}_{\sigma,\lambda=1}^\dagger$ associated with the eigenvalue 0, in agreement with the IFT $\langle e^{-\Sigma} \rangle = 1$. However, we stress that the two IFTs for Σ^m and Σ are valid at any time, and not only asymptotically.

2. Small- t expansion of $Z_{i^v}(\lambda, t)$

In order to investigate whether the symmetry of the SCGF $\mu_{i^v}(\lambda)$ [eq. (S28) above] reflects a more general symmetry of the pdf for the information flow, we introduce the modified function $\tilde{Z}_{i^v,\lambda}(v, y, t) = Z_{i^v,\lambda}(v, y, t) e^{\lambda \langle i^v \rangle t}$ which evolves according to

$$\partial_t \tilde{Z}_{i^v,\lambda}(v, y, t) = \tilde{\mathcal{L}}_{i^v,\lambda}^\dagger \tilde{Z}_{i^v,\lambda}(v, y, t) , \quad (\text{S47})$$

with $\tilde{\mathcal{L}}_{i^v,\lambda}^\dagger = \mathcal{L}_{i^v,\lambda}^\dagger + \lambda \langle i^v \rangle$. The solution can be formally expanded in powers of t as

$$\tilde{Z}_{i^v,\lambda}(v, y, t) = p(v, y) + \sum_{n=1}^{\infty} \frac{1}{n!} \tilde{Z}_{i^v,\lambda}^{(n)}(v, y) t^n \quad (\text{S48})$$

with $\tilde{Z}_{i^v,\lambda}^{(n)}(v, y) = (\tilde{\mathcal{L}}_{i^v,\lambda}^\dagger)^n p(v, y)$. For brevity, we here only report the expression of the term proportional to t ,

$$\tilde{Z}_{i^v,\lambda}^{(1)}(v, y) = p(v, y) [f_0(v, y) + f_1(v, y)\lambda + f_2(v, y)\lambda^2] , \quad (\text{S49})$$

with

$$\begin{aligned}
f_0(v, y) &= \frac{1}{2\tau^2 m^2 (c_{11}c_{22} - c_{12}^2)^2} \left[[2\gamma T \tau^2 c_{22}^2 + \Delta m^2 c_{12}^2 - 2m\tau(c_{11}c_{22} - c_{12}^2)(\gamma\tau c_{22} + mc_{12})]v^2 \right. \\
&\quad + [2\gamma T \tau^2 c_{12}^2 + \Delta m^2 c_{11}^2 + 2m\tau(c_{11}c_{22} - c_{12}^2)(\kappa\tau c_{12} - mc_{11})]y^2 \\
&\quad - 2[2\gamma T \tau^2 c_{12}c_{22} + \Delta m^2 c_{11}c_{12} - m\tau(c_{11}c_{22} - c_{12}^2)(c_{12}(\gamma\tau + m) + mc_{11} - \kappa\tau c_{22})]vy \\
&\quad \left. - (c_{11}c_{22} - c_{12}^2)(2\gamma T \tau^2 c_{22} + \Delta m^2 c_{11} - 2m\tau(c_{11}c_{22} - c_{12}^2)(\gamma\tau + m)) \right] \\
f_1(v, y) &= \frac{\gamma}{m^2 c_{11} (c_{11}c_{22} - c_{12}^2)^2} \left[[Tc_{22}(c_{11}c_{22} - 2c_{12}^2) - m(c_{11}c_{22} - c_{12}^2)^2]v^2 - Tc_{12}^2 c_{11}y^2 \right. \\
&\quad \left. + [2Tc_{12}^3 - \frac{\kappa m}{\gamma}(c_{11}c_{22} - c_{12}^2)^2]vy + Tc_{12}^2(c_{11}c_{22} - c_{12}^2) \right] \\
f_2(v, y) &= \frac{\gamma T c_{12}^2}{m^2 c_{11}^2 (c_{11}c_{22} - c_{12}^2)^2} (c_{12}v - c_{11}y)^2 .
\end{aligned} \tag{S50}$$

Remarkably, the term linear in λ cancels by integration over v and y , and we obtain

$$\tilde{Z}_{i^v}(\lambda, t) \equiv Z_{i^v}(\lambda, t)e^{\lambda \langle i^v \rangle t} = 1 + \frac{\gamma T}{m^2} \frac{c_{12}^2}{c_{11}(c_{11}c_{22} - c_{12}^2)} \lambda^2 t + \mathcal{O}(t^2) . \tag{S51}$$

The calculation of higher-order terms quickly becomes cumbersome and we have been able to perform the expansion of $\tilde{Z}_{i^v, \lambda}(v, y, t)$ up to order t^3 only. Again, we find, after integrating over v and y , that odd powers of λ do not contribute to $\tilde{Z}_{i^v}(\lambda, t)$. This leads us to conjecture that $\tilde{Z}_{i^v}(\lambda, t)$ is an even function of λ .

We have performed the same calculation for a non-linear model with a feedback force $f_t = -\kappa y_t + b y_t^3$, by expanding all functions in powers of the small parameter b . Results show that the symmetry is already lost at the first order in b and t .

B. Auxiliary processes

1. Biased forces

Once the right eigenfunction of the tilted generator associated with the dominant eigenvalue is computed, e.g. $r_{i^v, \lambda}(v, y) \propto e^{-(1/2)[A_{i^v}(\lambda)v^2 + B_{i^v}(\lambda)y^2 + 2C_{i^v}(\lambda)vy]}$ in the case of I^v , we can determine the auxiliary (or driven) process that realizes the conditioned process $\mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t | I^v = i^v t]$ in the limit $t \rightarrow \infty$. This process is a diffusion with the same two-dimensional matrix \mathbf{D}' as the original process and a modified drift given by (cf. eq. (19) in [26] with $k \rightarrow -\lambda$)

$$\mathbf{F}_{i^v, \lambda}(v, y) = \mathbf{F}(v, y) - \mathbf{D}'[\lambda \mathbf{g}_{i^v}(v, y) - \nabla \ln r_{i^v, \lambda}(v, y)] . \tag{S52}$$

Namely, the auxiliary dynamics is governed by the coupled linear equations [eqs. (30) in the main text]

$$\begin{aligned}
m\dot{v}_t &= -\gamma_{\text{eff}} v_t - \kappa_{\text{eff}} y_t + \xi_t \\
\tau \dot{y}_t &= k_1 y_t + k_2 v_t + \eta_t .
\end{aligned} \tag{S53}$$

with

$$\begin{aligned}
\frac{\gamma_{\text{eff}, i^v}}{\gamma} &= 1 + \frac{2\lambda T}{m}(\alpha_{11} - c_{11}^{-1}) + \frac{2T}{m} A_{i^v}(\lambda) , \quad \frac{\kappa_{\text{eff}, i^v}}{\kappa} = 1 + \frac{2T}{m} \frac{\gamma}{\kappa} [\lambda \alpha_{12} + C_{i^v}(\lambda)] \\
k_{1, i^v} &= -1 - \frac{\Delta}{\tau} B_{i^v}(\lambda) , \quad k_{2, i^v} = 1 - \frac{\Delta}{\tau} C_{i^v}(\lambda) .
\end{aligned} \tag{S54}$$

Likewise, the coefficients of the effective drifts for Σ^m and Σ are obtained as

$$\begin{aligned}
\frac{\gamma_{\text{eff}, \sigma^m}}{\gamma} &= 1 - 2\lambda + \frac{2T}{m} A_{\sigma^m}(\lambda) , \quad \frac{\kappa_{\text{eff}, \sigma^m}}{\kappa} = 1 + \frac{2T}{m} \frac{\gamma}{\kappa} C_{\sigma^m}(\lambda) \\
k_{1, \sigma^m} &= -1 - \frac{\Delta}{\tau} B_{\sigma^m}(\lambda) , \quad k_{2, \sigma^m} = 1 - \frac{\Delta}{\tau} C_{\sigma^m}(\lambda) ,
\end{aligned} \tag{S55}$$

and

$$\begin{aligned} \frac{\gamma_{\text{eff},\sigma}}{\gamma} &= 1 + \frac{2\lambda T}{m}(\alpha_{11} - \frac{m}{T}) + \frac{2T}{m}A_\sigma(\lambda) \quad , \quad \frac{\kappa_{\text{eff},\sigma}}{\kappa} = 1 + \frac{2T}{m} \frac{\gamma}{\kappa} [\lambda\alpha_{12} + C_\sigma(\lambda)] \\ k_{1,\sigma} &= -1 - \frac{\Delta}{\tau}B_\sigma(\lambda) \quad , \quad k_{2,\sigma} = 1 - \frac{\Delta}{\tau}C_\sigma(\lambda) \quad , \end{aligned} \quad (\text{S56})$$

respectively.

2. Biased noises

As briefly explained in the main text, thanks to the linearity of the model, we can also define atypical Gaussian noises ξ_{atyp} and η_{atyp} that create rare fluctuations of σ^m or v^v . In the frequency domain, a solution of eqs. (S53) reads

$$\begin{aligned} v_{\text{atyp}}(\omega) &= \chi_{\text{eff}}(\omega) \left[\xi(\omega) + \frac{\kappa_{\text{eff}}}{k_1 + i\omega\tau} \eta(\omega) \right] \\ y_{\text{atyp}}(\omega) &= -\frac{\chi_{\text{eff}}(\omega)}{k_1 + i\omega\tau} [k_2 \xi(\omega) + (\gamma_{\text{eff}} - im\omega) \eta(\omega)] \end{aligned} \quad (\text{S57})$$

where $\chi_{\text{eff}}(\omega)$ is given by eq. (32) in the main text. By inserting this solution into the original equations of motion, we then express the two biased noises in terms of the original Gaussian white noises (eqs. (31) in the main text). Their power spectral densities are given by

$$\begin{aligned} S_{\xi_{\text{atyp}}}(\omega) &\equiv \langle \xi_{\text{atyp}}(\omega) \xi_{\text{atyp}}(-\omega) \rangle = \frac{|\chi_{\text{eff}}(\omega)|^2}{k_1^2 + \omega^2 \tau^2} \left[2\gamma T [(\gamma k_1 - \kappa k_2 + m\omega^2 \tau)^2 + \omega^2 (\gamma \tau - m k_1)^2] \right. \\ &\quad \left. + \Delta [(\gamma \kappa_{\text{eff}} - \gamma_{\text{eff}} \kappa)^2 + m^2 \omega^2 (\kappa - \kappa_{\text{eff}})^2] \right] \\ S_{\eta_{\text{atyp}}}(\omega) &\equiv \langle \eta_{\text{atyp}}(\omega) \eta_{\text{atyp}}(-\omega) \rangle = \frac{|\chi_{\text{eff}}(\omega)|^2}{k_1^2 + \omega^2 \tau^2} \left[2\gamma T [(k_1 + k_2)^2 + \omega^2 \tau^2 (1 - k_2)^2] \right. \\ &\quad \left. + \Delta [(\gamma_{\text{eff}} + \kappa_{\text{eff}} - m\omega^2 \tau)^2 + \omega^2 (m + \gamma_{\text{eff}} \tau)^2] \right] . \end{aligned} \quad (\text{S58})$$

The variations of the corresponding intensities $S_{\xi_{\text{atyp}}}(\omega = 0)$ and $S_{\eta_{\text{atyp}}}(\omega = 0)$ as a function of σ^m and v^v are shown in Fig. 7 of the main text.

3. Relationship with the modified dynamics and the IFTs

Finally, we uncover the relationship between the auxiliary (or driven) dynamics and the modified dynamics introduced above in section A to derive the IFTs. As an example, we consider the so-called “star” dynamics associated with the IFT $\langle e^{-\Sigma} \rangle = 1$ and defined by eq. (S5). Comparing eq. (S1) with the equation defining the exponentially tilted path ensemble for Σ

$$\mathcal{P}_{\sigma,\lambda}[\mathbf{v}_0^t, \mathbf{y}_0^t] \equiv \frac{e^{-\lambda \Sigma} \mathcal{P}[\mathbf{v}_0^t, \mathbf{y}_0^t]}{\langle e^{-\lambda \Sigma} \rangle} \quad , \quad (\text{S59})$$

we readily see that

$$\mathcal{P}^*[\mathbf{v}_0^t, \mathbf{y}_0^t] = \mathcal{P}_{\sigma,\lambda=1}[\tilde{\mathbf{v}}_0^t, \tilde{\mathbf{y}}_0^t] \quad (\text{S60})$$

for any trajectory of duration t . Since the path measure of the auxiliary process becomes equivalent to the tilted path measure as $t \rightarrow \infty$, we thus conclude that the probability to observe a trajectory with the star process and the probability to observe the time-reversed trajectory with the auxiliary process are asymptotically identical when $\lambda = 1$. In particular, the corresponding stationary pdfs are simply related by time reversal

$$p^*(v, y) = p_{\sigma,\lambda=1}(-v, y) \quad . \quad (\text{S61})$$

In other words, the star process and the time reversal of the auxiliary process for $\lambda = 1$ must be governed by the same equations of motion in the stationary limit. To check this identity explicitly, we build the time reversal

of the auxiliary process, carefully taking into account the fact that v_t is an odd variable. Since the process is a two-dimensional Ornstein-Uhlenbeck process, its time reversal is again a diffusion governed by the coupled equations

$$\dot{\mathbf{X}}_t = -\mathbf{F}_{\sigma,\lambda}(\mathbf{X}_t) - \mathbf{D}'\mathbf{C}_{\sigma,\lambda}^{-1}\mathbf{X}_t + \boldsymbol{\xi}_t, \quad (\text{S62})$$

where $\mathbf{X}_t = (-v_t, y_t)$, $\mathbf{F}_{\sigma,\lambda}(\mathbf{X}_t)$ is the two-dimensional force corresponding to the effective drifts defined by eqs. (S56), and $\mathbf{C}_{\sigma,\lambda}$ is the stationary covariance matrix in the auxiliary process, hence $p_{\sigma,\lambda}(v, y) = l_{\sigma,\lambda}(v, y)r_{\sigma,\lambda}(v, y) \propto e^{-(1/2)\mathbf{X}^T\mathbf{C}_{\sigma,\lambda}^{-1}\mathbf{X}}$. We then set $\lambda = 1$ in these equations and use the fact that $l_{\sigma,\lambda=1}(v, y) = p(v, y)$ (see the remark at the end of section C.1.a). Therefore,

$$p_{\sigma,\lambda=1}(v, y) \propto e^{-\frac{1}{2}\left\{[\alpha_{11}+A_\sigma(1)]v^2 + [\alpha_{22}+B_\sigma(1)]y^2 + 2[\alpha_{12}+C_\sigma(1)]yv\right\}}, \quad (\text{S63})$$

which gives the expression of $\mathbf{C}_{\sigma,\lambda=1}^{-1}$. Inserting into eqs. (S62), we find that the terms involving the quantities $A_\sigma(1)$, $B_\sigma(1)$ and $C_\sigma(1)$ cancel out and we finally recover the equations of motion of the star process.

We stress that the asymptotic equivalence between the two processes is only valid when the right and left eigenfunctions $r_{\sigma,\lambda=1}(v, y)$ and $l_{\sigma,\lambda=1}(v, y)$ are normalizable and the pre-exponential factor $g_\sigma(\lambda = 1)$ is finite. This latter condition is not satisfied when the stationary state of the star process does not exist (i.e. $\int dv dy p^*(v, y)$ diverges). Then eq. (S61) does not hold and $l_{\sigma,\lambda=1}(v, y) \neq p(v, y)$. However, it is noteworthy that the stationary pdf of the auxiliary process $p_{\sigma,\lambda=1}(v, y) = l_{\sigma,\lambda=1}(v, y)r_{\sigma,\lambda=1}(v, y)$ is still normalizable.